Computers & Education 151 (2020) 103862

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

Chatbots for learning: A review of educational chatbots for the Facebook Messenger

Pavel Smutny *, Petra Schreiberova

VSB – Technical University of Ostrava, Czech Republic

ARTICLE INFO

Keywords:
Chatbot
Messenger
Facebook
Mobile learning
Quality evaluation

ABSTRACT

With the exponential growth in the mobile device market over the last decade, chatbots are becoming an increasingly popular option to interact with users, and their popularity and adoption are rapidly spreading. These mobile devices change the way we communicate and allow ever-present learning in various environments. This study examined educational chatbots for Facebook Messenger to support learning. The independent web directory was screened to assess chatbots for this study resulting in the identification of 89 unique chatbots. Each chatbot was classified by language, subject matter and developer's platform. Finally, we evaluated 47 educational chatbots using the Facebook Messenger platform based on the analytic hierarchy process against the quality attributes of teaching, humanity, affect, and accessibility. We found that educational chatbots on the Facebook Messenger platform vary from the basic level of sending personalized messages to recommending learning content. Results show that chatbots which are part of the instant messaging application are still in its early stages to become artificial intelligence teaching assistants. The findings provide tips for teachers to integrate chatbots into classroom practice and advice what types of chatbots they can try out.

1. Introduction

A chatbot is a software tool that interacts with users on a certain topic or in a specific domain in a natural, conversational way using text and voice. For many different purposes, chatbots have been used across a wide range of domains, including marketing, customer service, technical support, as well as education and training. Current developments in this area suggest that interaction with technologies, either by natural language or by speech, is possible because technology develops, and users become more used to interacting with digital entities. Rather than creating a human-like smart machine application, it is about creating effective digital assistants who are able to provide information, answer questions, discuss a specific topic, or perform a task.

Personal digital assistants like Siri from Apple, Alexa from Amazon, Microsoft's Cortana or Assistant from Google are at the forefront of technology of voice recognition and artificial intelligence. These digital assistants use machine-learning techniques and are able to manage some day-to-day tasks of traditional assistants or secretaries (such as email prioritization, highlighting the most important content and interactions) to help their users become more effective. A vast number of simpler and more domain-specific text-based chatbots complement target-specific functionalities such as raising support tickets to leave feedback, disseminating content for publishing sites, booking a hotel room, making a restaurant reservation, etc. Text-based chatbots typically follow a set of established rules or flows to respond to questions posted by a user (Budiu, 2018). These rules or flows enable them to respond

* Corresponding author.
E-mail addresses: pavel.smutny@vsb.cz (P. Smutny), petra.schreiberova@vsb.cz (P. Schreiberova).

https://doi.org/10.1016/j.compedu.2020.103862
Received 21 June 2019; Received in revised form 21 February 2020; Accepted 27 February 2020
Available online 29 February 2020

effectively to requests within a specific domain, but are not efficient in answering questions, the pattern of which does not match the rules on which the chatbot is trained.

Chatbot applications have been around for a long time, notable examples being ELIZA, ALICE, Claude, and HeX. ELIZA was the world's very first chatbot. Developed by Joseph Weizenbaum in 1956 (Weizenbaum, 1983), it was designed to emulate a psychotherapist and had a knowledge base in this domain. ALICE (Artificial Linguistic Internet Computer Entity) was developed by Richard S. Wallace as a chatbot based on natural language processing. It uses Artificial Intelligence Markup Language (AIML) (Shawar & Atwell, 2015), which is responsible for pattern matching and to relate a user input with a response of more than 40,000 knowledge records. Other current chatbots nowadays use the framework of ALICE. Brian McLaughlin developed a chatbot called Claude that uses standard pattern matching to find a suitable answer (Laven, 1996). The Claude chatbot recognizes user input, then develops a response based on that input using responses in its database, and finally constructs an answer. Jason Hutchens (Hutchens, 1997) developed the HeX chatbot, that was able to give a reply and also introduce a new topic with a certain probability. (Wu, Wang, Li, & Li, 2008).

Chatbots have had a long history of use as pedagogical agents in educational settings. From the early 1970s, pedagogical agents within digital learning environments known as Intelligent Tutoring Systems have been developed (Laurillard, 2013). Conversational pedagogical agents use artificial intelligence techniques to enhance and personalize automation in teaching. The design and research knowledge are important in developing engaging, useful, and valuable pedagogical agents that not only make the most of technological advancements, but also understand emotional, cognitive, and social educational concerns (Gulz, Haake, Silvervarg, Sjödén, & Veletsianos, 2011) (King, 2002). In addition, conversational agents have been built into software and devices. Moreover, in recent years more and more organizations have started to exploit their capabilities beyond simple querying of information followed by a programmed response. However, understandings of these innovations are frequently untheorized and immature. Incorporation of chatbots into the educational area over the last decade implies an increase in interest in the ways in which chatbots might be implemented for teaching and learning. Useful chatbot systems can provide benefits of instant availability and ability to respond naturally through a conversational interface with the same advantages as an interview. Additionally, chatbots demonstrate the ability to create easygoing interactions with users so that they can be leveraged to support engagement, as well as setting out goals, strategies and outcomes of learning and training (Cinglevue, 2017).

Over the last few years, along with the attractiveness of instant messaging, chatbots and pedagogical agents have motivated educators to integrate messaging tools in teaching and learning. Coniam (Coniam, 2014) evaluated five renowned language chatbots concluding that they do not yet make robust chatting partners but have significantly improved, with three of the five performing satisfactorily at the level of grammar. Chatbot Ethnobot (Tallyn, Fried, Gianni, Isard, & Speed, 2018) replaces human ethnographer, collects ethnographic data and asks participants a series of questions in a chat format. Chatbot technology also shows potential as a teaching and learning tool in the distance and online education (Heller, Proctor, Mah, Jewell, & Cheung, 2005). Two educational chatbots built on Facebook Messenger platform helps students with introductory and intermediate accounting area. The majority (72% of the users) expressed their overall satisfaction with these chatbots as virtual tutors (Schmulian & Coetzee,). The aim of the study by Crutzen and colleagues (Crutzen, Peters, Portugal, Fisser, & Grolleman, 2011) was to look into how adolescents used a chatbot that answers questions about sex, drugs, and alcohol, particularly in contrast with information phone lines and search engines.

Mobile devices offer a wide range of ways to work together, communicate and learn. However, young generations do not differentiate between tools and mobile devices for social networking. The advantages of mobile devices were identified as access to multimedia content, portability, flexibility and the ability to search for information immediately (Gikas & Grant, 2013) (Sung, Chang, & Liu, 2016). The disadvantages of mobile devices include reading learning content on a small screen, lack of focus and effectiveness of attention, technological issues (battery life, connectivity) or device compatibility issues (Chartrand, 2016, pp. 1–13). Research in the field of mobile instant messaging have focused on the use of a mobile instant messaging tool WhatsApp to support teaching and learning (So, 2016) (Pimmer et al., 2019). The participants of the experiment showed positive perception and acceptance of the use of the mobile instant messaging tool for teaching and learning. Also, they did not really mind receiving instructional materials and questions outside school hours. The mobile devices allow students to remain engaged in formative guidance and receive feedback, which is required in encouraging a learner-centred environment. (Roblyer, McDaniel, Webb, Herman, & Witty, 2010). Research supports the potential for student-to-student and student-to-instructor interactions with the use of Facebook social networking. The research by (Yildirim, 2017) resolves that teaching practices based on gamification enhance student attitudes towards lessons and have a positive influence on student achievement. Meanwhile, research by (Westerman, Daniel, & Bowman, 2016) (Awidi, Paynter, & Vujosevic, 2019) examined positive attitudes of students toward social media.

Today, the chatbot landscape is wide. Chatbots are not associated as a single category but they fall along a wider spectrum. We propose the following classification by input or by messaging channels.

By input.

- *Button-Based* Decision tree hierarchies are presented to the user in the form of buttons.
- *Keyword Recognition-Based* User types in a phrase or just a word and the chatbot matches the input with a pre-loaded response.
- *Contextual* Utilize machine learning and artificial intelligence to self-improve based on what users are asking for and how they are asking it.
- *Voice-Enabled* User inputs through voice and chatbot using voice recognition answers user's queries or performs creative tasks.

By messaging channels.

- Standalone application

- Desktop (example: Braina Virtual Assistant)
- mobile (examples: applications Andy English, DoNotPay or Replika: My AI Friend)

- Web-based service
  - integrated on the web (examples: popup window with customer service help)
  - individual (examples: Mitsuku, Cleverbot)
- Integrated
  - instant messaging apps (examples: Facebook Messenger, WhatsApp, WeChat, Skype)
  - communication and collaboration platform (examples: Slack, Microsoft Teams, Cisco Webex Teams)

Social media and networking service Facebook opened up its Messenger platform to let chatbots into the application in April 2016. Since then, the Facebook Messenger platform has over 1.3 billion monthly users (Constine, 2018), accounts with over 300,000 chatbots, as well as businesses and customers that exchange 8 billion messages a day (Johnson, 2018). The major advantage of using a Facebook Messenger chatbot is a low barrier to entry for the creator and his target audience. Facebook Messenger is the third most-used mobile application in the world, used by 68 percent of users (Hartmans, 2017). Benefits for the users are using a familiar interface, no need to download and install extra application, 24/7 availability. Drawback includes interference from other conversations in the instant messaging application (Pereira & Díaz, 2018). There are several advantages for the chatbot creator: using the existing infrastructure of social platform, personalization of the conversation, presence on mobile and web. Chatbots can be developed in any programming language or use a third-party no-code platform that uses a visual development environment. Issues with chatbots are that users are not familiar with a chatbot experience, poor mechanisms for discoverability and limitations based on policy and usage guidelines.

The purpose of this study was to explore the use of a chatbot in the instant messaging tool to support learning. The two following research questions were addressed in this study:

Research Question 1: What is the distribution of subject matter, use of language and development environment in chatbots on the Facebook Messenger platform?

Research Question 2: What is the quality of educational chatbots on the Facebook Messenger platform?

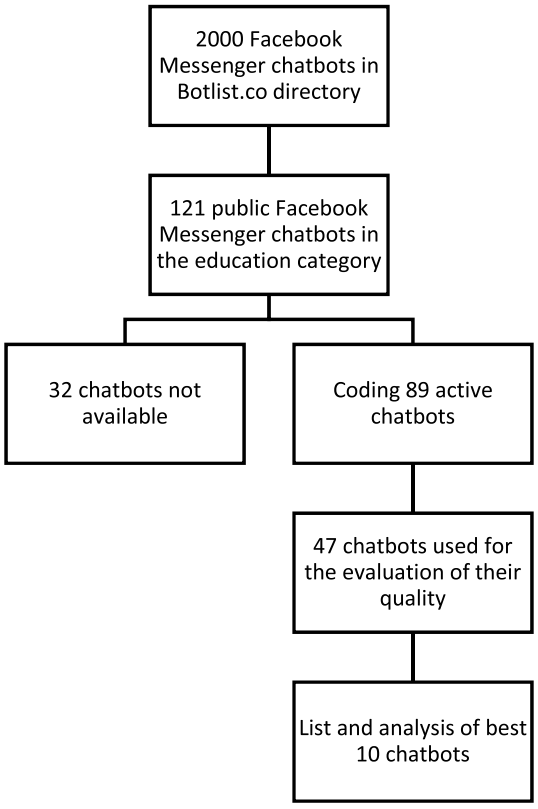

Fig. 1. Search, screening and selection process for educational chatbots.

2. Methods

2.1. Search strategy

Compared with applications stores like Apple's App Store, Microsoft Store or Google Play there is no single space where users can filter or discover educational chatbots. If we want to find a chatbot directly on the Facebook Messenger, we need to search the name of the brand or chatbot in the search bar. However, users would rather look for a specific service than the exact name or brand. There are independent web directories like Chatbottle (chatbottle.co) or Botlist (botlist.co) that find bots. Botlist directories list chatbots from various platforms (e.g. Amazon Echo, Skype, Slack, and others), with over 2000 Messenger chatbots in 29 categories.

Screening and assessment of chatbots for study inclusion were guided by a systematic review process and restricted to the Botlist directory (Fig. 1). The directory was inventoried during the period of March–April 2018, resulting in the identification of 121 total chatbots (N = 121). Chatbots meeting the following inclusion criteria were analyzed further: (a) education category, (b) available to the public, (c) chatbot for Facebook Messenger. However, 32 bots were found inactive or not available, resulting in 89 unique chatbots (N = 89).

2.2. Coding of chatbots

A coding instrument was developed to identify chatbots in the following categories: language, subject matter, and platform (Table 1). Each chatbot was first classified by *Language,* in which the bot communicated with the user. The majority of the chatbots used English as a communication language (89%). Other languages included French (N = 8, 9%) and Arabic (N = 7, 8%). Altogether, 11 languages were identified through conversation with the chatbots, with a group of chatbots being designed to communicate with more than one language (N = 9, 10%). Each chatbot was then classified by *Subject matter.* Within this classification, twelve areas were identified. The majority of the bots were classified as "Information" (N = 42, 47%), and their educational content was questionable. These chatbots often only use automated answers or button-based browsing, which lead to website links outside the chatbot interface, and lack any discussion techniques or simulation of the human conversation. Therefore only chatbots classified as Language, Economics, Math, Multiple subject matter, Religion, Literature, History, Nature, Programming, Psychology, or Design will be used (N = 47, 53%) for the evaluation of their quality. Finally, each chatbot was classified by *Platform.* There are two options on how to create a chatbot on Facebook Messenger platform: with coding using the Messenger Platform and sets of APIs, or without coding using chatbot building tools that offer the ability to set up a chatbot and connect it to a Facebook Page. Over 55% of the chatbots used a platform

Table 1
Characteristics of educational chatbots included in the study (N = 89).

	N	%
Language		
English	79	89
French	8	9
Arabic	7	8
Italian	3	3
Russian	3	3
Spanish	2	2
Korean	2	2
Bengali	1	1
German	1	1
Japanese	1	1
Hindi	1	1
Subject matter		
Information	42	47
Language	16	18
Economics	6	7
Multiple subject matter	5	6
Math	4	5
Religion	4	5
Literature	3	3
History	2	2
Nature	2	2
Programming	2	2
Psychology	2	2
Design	1	1
Platform		
Chatfuel	36	40
n/a	49	55
Other (Adexin, ManyChat, Motion.ai, Custombotdesign.com)	4	5

which could not be identified from the user's perception. However, the Chatfuel service was identified as the most common platform (N = 36, 40%) used between building tools.

2.3. Evaluating the quality of chatbots

In general, when having a conversation with a chatbot, users expect that chatbots will be able to answer questions and process orders, whilst also allowing for the automation of certain types of routine, repetitive, time-consuming communication. Interactions with a chatbot in Messenger tend to be more natural than with a mobile application or a website. People tend to ask the question they have in mind and expect to get an answer. Yet Weinberg (Weinberg, 2017) reports Facebook has seen 70% of its chatbot interactions fail, meaning that the artificial intelligence could not understand what users were saying and humans had to step in, according to a news report.

Radziwill and Benton (Radziwill & Benton, 2017) proposed a quality assessment method for evaluating the quality of chatbots and intelligent conversational agents. Based on literature reviews of quality issues and attributes, they recommend a method based on the Analytic Hierarchy Process (AHP). This approach is suitable for solving multicriteria decision-making problems. Radziwill and Benton summarized the quality attributes from previous studies as a checklist for chatbot analysis. The quality attributes include the effectiveness (functionality, humanity), efficiency (performance), and satisfaction (accessibility, affect, behavior and ethics).

For our study, we used this checklist to limit the subjective factor of the decision-maker. Based on the objective and the theme of our study, we considered four main categories for the decision using AHP (Teaching, Humanity, Affect, Accessibility) and appropriate quality attributes (Table 2).

The Analytic Hierarchy Process (AHP) is a structured technique for organizing and analyzing complex decisions that involve both qualitative and quantitative considerations. The AHP is a group decision-making method originally developed by prof. Thomas L. Saaty (Saaty, The Analytic Hierarchy Process, Saaty, 1980). It is a method for ranking and prioritizing alternatives where it is necessary to consider multiple criteria (Saaty, Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process, Saaty, 2008). Firstly, one decomposes their decision problem into a hierarchy of quality attributes and select metrics. Secondly, the decision makers evaluate various elements by comparing them to each other two at a time, with respect to their impact on an element above them in the hierarchy. Finally, numerical priorities are calculated for each of the decision alternatives to determine which option best satisfies the hierarchy of quality attributes.

3. Results

The Analytic Hierarchy Process (Saaty, How to make a decision: The analytic hierarchy process, Saaty, 1990) can be described as a method of decomposition of unstructured situations into a simpler one. This method assigns a value to each variant based on the use of the subjective evaluation of the pairwise comparison. Subsequent synthesis identifies the alternative with the highest priority. The aim of the authors is to include into the model as much relevant information as possible, own goals, criteria, and variants and to determine the optimal solution.

The target of our research is to compare the Facebook Messenger chatbots suitable for learning in the selected areas. The methodology of the chatbots selection is described in Sections 2.2 and 2.3. The selected 47 chatbots were defined as variants of our model.

We assigned a rating to each variant (47 chatbots) for all formulated quality attributes (Table 2). The evaluation was conducted as a chat conversation: a single evaluator rated in a 1–100 scale answers by the chatbot. The evaluator had scripted questions for the chatbots, but he had to adapt them during the conversation according to the topic or chatbot style of dialogue (e.g., whether the chatbot was able to answer open questions or just used predefined navigation buttons). The hierarchical structure of the model is shown in Table 3.

The next step is to construct a set of pairwise comparison matrices between all criteria and subcriteria. We assign relative ranking to each pair of criteria and subcriteria to determine priorities. An approach based on the Saaty scale was used to compare the criteria

Table 2
Attributes for chatbot quality assessment.

Category	Quality Attribute	Examples
Teaching	Recommends learning content	Links to web pages or documents with learning topics.
	Provides feedback, Q&A	Quizzes and tests provide instant feedback on each question at the time that question is answered.
	Set goals and monitor learning progress	Quizzes and tests are counting scores. Ability to resume test/quiz and continue later.
Humanity	Able to maintain themed discussion	Chatbot interprets commands accurately. Natural and convincing conversation.
	Able to respond to specific questions	Flexible interpretation of knowledge.
Affect	Provides greetings, pleasant personality	Says Hello. Greet the human participant with its name.
	Entertaining, engaging	Chatbot uses jokes, humor, emoji, animated GIFs from popular culture.
Accessibility	Can detect meaning and intent	Response to the question "Sorry, I don't speak English. "or "Sorry, could you repeat the question please?"
	Responds to social cues appropriately	Read and respond to the moods of the human participant.

Table 3
Hierarchical structure.

Goal	Criterion	Subcriterion	Variant
Chatbot	Teaching	Feedback	Chatbot1
		Progress	Chatbot2
		LearningContent	.
	Humanity	SpecificQs	.
		Themed Discussion	.
	Affect	Entertaining	.
		Personality	.
	Accessibility	Meaning Intent	.
		Social Cues	Chatbot47

(Saaty, The Analytic Hierarchy Process, Saaty, 1980). The relative importance between two criteria is measured according to a numerical scale, as shown in Table 4. If the criterion is more important than the second one, we assign it the value $s_{ij} \in \{1, 3, 5, 7, 9\}$ otherwise $s_{ij} = \frac{1}{s_{ji}}$.

By this mutual comparison, we constructed the matrix of pairwise comparisons. Table 5 shows the final comparison matrix for criteria where each cell indicates how much more important the criteria is in the row in comparison to the criteria in the column. For example, the comparison between Humanity and Teaching – Teaching is preferred strongly (5 times) over Humanity, so the value 5 is entered in the position (1, 2) and the reciprocal value 1/5 is automatically entered in the transpose position (2, 1).

From the matrix, we determine the principal eigenvalue, eigenvector and by its normalization we get the required weights (the scale of priorities). The computed priority vectors represent the relative weights among the criteria that we compared. One can see (Fig. 2) that Teaching is 57%, Humanity is 23%, Affect is 15%, and Accessibility only 5% for educational chatbots. It is visible that the criterion Teaching is the most important for our study. We can say that Teaching is 2.48 times more important than Humanity (3.8 times more important than Affect and 11.4 times more than Accessibility).

The same procedure was applied for all subcriteria. Based on the calculated weights, we get an overall evaluation and ranking of the variants according to defined priorities.

Having analyzed 47 chatbots using human-chatbot conversations, we concluded that the model is too complicated and is not appropriate for the relevant decision. Therefore, we modified the model so that we chose 10 chatbots, which obtained the highest score from the all 47. The model for the final analysis is shown in Fig. 3.

The results of the analysis are shown in Table 6 and Table 7. According to the final evaluation, the AskFrank variant with the total weight of 14.9 was the best placed, which represents an optimal decision based on the model.

The influence of each criterion on the top ten chatbots is shown in Fig. 4. Illustrating the priority alternatives, it details the contribution of each criterion in the final priority score. Obviously, the influences of the criteria are different from one another. The criterion of Humanity was a very important ability for users using the chatbots that took the top two places (*AskFrank* and *IFRSRookies*). The *Wordsworth* and *EnglishWithEdwin* chatbots were better suited to users who mainly required teaching properties (setting goals, monitoring the learning process, feedback, etc.). Only two of the top ten chatbots (*AskFrank* and *IFRSRookies*) were primarily focused on the criterion of Humanity (able to respond to specific questions, able to maintain themed discussion).

The value of the consistency ratio is close to 10%. It means that our evaluation of chatbot preference should be consistent and acceptable. One way we can achieve better consistency is a deletion of Accessibility property or to reassess the subjective judgment.

The AHP package in the R Statistical Software (R Core Team, 2014) was used to define and evaluate the model. The advantage of this application is the very easy modification of the values of the criteria and therefore it is possible to analyze the behavior of the solution due to input changes by means of repeated simulations.

4. Discussion

The purpose of the present research was to investigate the current availability of educational chatbots on the Facebook Messenger platform and examine their quality. Although web-based chatbots have been available for several decades, a big step forward comes from chatbots moving to an instant messaging environment. Beyond texting, an instant messaging application can include additional features and capabilities: voice and video messages, calling, gaming, e-commerce, media sharing, animated GIFs, stickers, advertisements or chatbots. Downloading and installing a single-purpose application is not necessary. Using an instant messaging platform

Table 4
Saaty scale.

Scale	Degree of preference
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance

Table 5
Final reciprocal matrix for pairwise comparisons.

	Teaching	Humanity	Affect	Accessibility
Teaching	1	5	3	7
Humanity	1/5	1	3	5
Affect	1/5	1/3	1	5
Accessibility	1/7	1/5	1/5	1

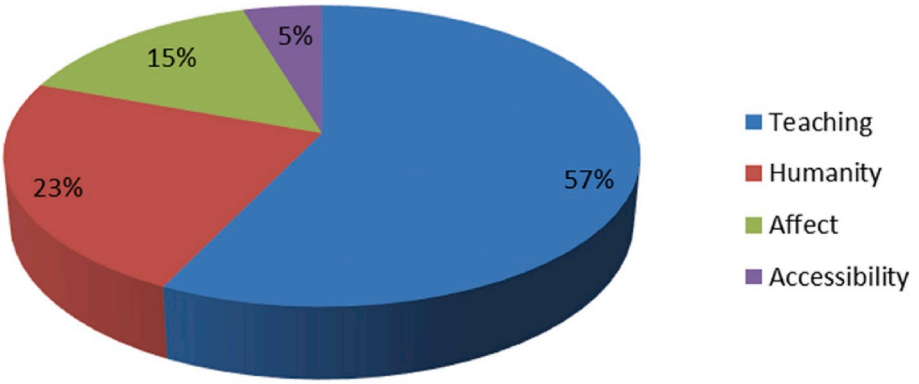

Fig. 2. Preference of criterions.

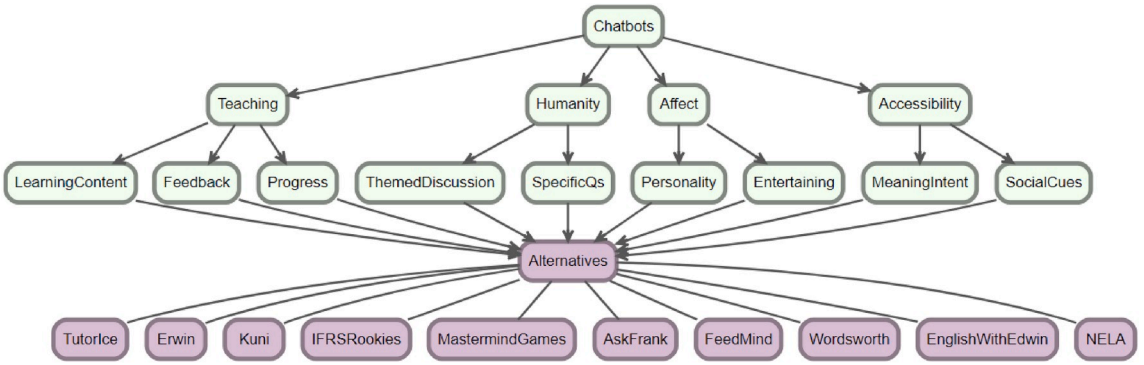

Fig. 3. Final hierarchical structure for 10 chatbots.

such as Facebook Messenger for hosting a chatbot brings the benefits of a familiar user interface and utilization of existing mobile infrastructure.

This research was the first to review educational chatbots on the Facebook Messenger platform and examines them based on the subject matter, conversation language and development platform. The findings from our research show that the majority of the chatbots (89%) used English as a communication language. Results also showed, nearly half of chatbots (46%) were lacking any discussion techniques or simulation of the human conversation and used button-based browsing or automated answers with additional information on websites outside the chatbot interface. Based on subject matter analysis, chatbots mostly deal with learning languages, economic topics or multiple subject matter. From the developer's point-of-view, the Chatfuel service was identified as the most common platform (40%) used between building tools.

Unfortunately, over (N = 23, 26%) of these chatbots give the impression of being short-lived – they were no longer available by the end of our study. Therefore, it remains difficult for teachers to provide guidance in terms of specific chatbots a student might consider an experiment with. However, findings of this content analysis do at least provide a general idea of the types of chatbots a student or teacher might take into account.

With the exponential growth of chatbots, we can ask how intelligent chatbots are? According to the report (Weinberg, 2017), 70% of Facebook Messenger chatbots were failing to fulfil user requests. Therefore, rather than building a large chatbots ecosystem, Facebook started to train Messenger chatbots to deal with a narrower set of cases so users would not be disappointed by the automation limitations. Next step for chatbots was recent progress in natural language processing and machine learning. They are constantly learning and improving from conversations with users and do generate responses based on collections of known conversations saved in

Table 6
Top ten chatbots based on evaluation using the AHP package in the R Statistical Software.

	Weight	AskFrank	IFRSRookies	Wordsworth	EnglishWithEdwin	FeedMind	Erwin	Kuni	NELA	MastermindGames	TutorIce
Chatbots	100.0%	14.9%	12.5%	11.0%	10.4%	10.1%	8.9%	8.7%	8.4%	8.0%	7.0%
Teaching	57.2%	3.4%	4.0%	8.2%	8.8%	6.1%	5.5%	5.0%	6.8%	4.4%	5.0%
Feedback	26.7%	2.9%	2.3%	2.9%	2.9%	2.9%	2.9%	2.9%	1.5%	2.3%	2.9%
Progress	26.7%	0.3%	0.3%	5.2%	5.2%	2.0%	2.4%	2.0%	5.2%	2.0%	2.0%
LearningContent	3.8%	0.1%	1.3%	0.1%	0.7%	1.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Humanity	23.4%	9.1%	5.8%	1.0%	1.0%	1.0%	1.4%	1.0%	1.0%	1.0%	1.3%
SpecificQs	19.5%	7.8	5.1%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%
ThemedDiscussion	3.9%	1.3%	0.6%	0.1%	0.1%	0.1%	0.6%	0.1%	0.1%	0.1%	0.5%
Affect	14.8%	0.4%	2.5%	1.5%	0.4%	2.8%	1.7%	2.5%	0.4%	2.3%	0.4%
Entertaining	12.3%	0.2%	2.3%	1.3%	0.2%	2.6%	1.3%	2.0%	0.2%	2.0%	0.2%
Personality	2.5%	0.2%	0.2%	0.2%	0.2%	0.2%	0.4%	0.5%	0.2%	0.4%	0.2%
Accessibility	4.7%	2.1%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%
MeaningIntent	4.1%	2.0%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%
SocialICues	0.6%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%

Table 7
Description of top ten chatbots based on evaluation.

Rank	Name	Percent	Subject matter	Language	Description
1.	Ask Frank	14.9	Multiple	English	Chatbot Frank can help answer any simple questions in mathematics, science and history.
2.	IFRSRookies	12.5	Economics	English	Chatbot teaches International Financial Reporting Standards at the introductory or intermediate level.
3.	Wordsworth	11.0	Language	English	Users can sign up for a daily lesson or take a fun quiz to test their vocabulary skills.
4.	English With Edwin	10.4	Language	English Spanish Korean Japanese Hindi Russian Arabic	Chatbot Edwin.ai is an online English-language learning tutor power by artificial intelligence.
5.	Feed.Mind	10.1	Multiple	French	Feed.Mind helps strengthen each week knowledge on a theme from economy, history, geopolitics, culture, politics, science, sport or technology
6.	Erwin	8.9	Math	English	Chatbot Erwin focuses on logic and presents a selection of sophisticated riddles and clues.
7.	Kuni	8.7	Economics	English	Chatbot Kuni tests Social Media Skills in a fun and interactive way.
8.	NELA	8.4	Language	English	Chatbot NELA helps improve language skills using word chain game.
9.	Mastermind Games	8.0	Psychology	English French Arabic	Chatbot explores users‘ mind through mastermind games.
10.	TutorIce	7.0	Math	English	Chatbot TutorIce helps learners by providing them with personalized and engaging math practice questions.

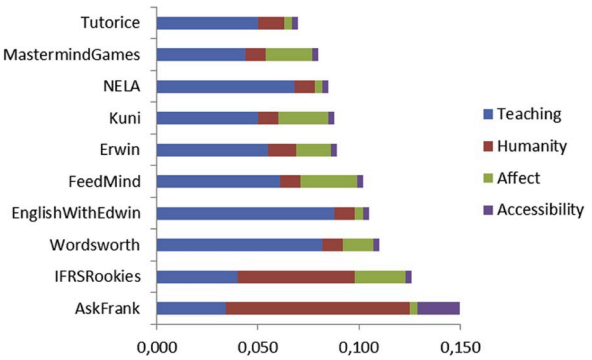

Fig. 4. Top ten chatbots comparison.

databases. This is perhaps an appropriate occasion to raise the question, if chatbots are built to process a large amount of data, why not use them as teaching assistants? In 2016, Professor Ashok K. Goel developed chatbot called Jill Watson for students enrolled in a Georgia Institute of Technology's Master program (Eicher, Polepeddi, & Goel, 2018). This virtual teaching assistant is based on IBM's Watson and is able to answer routine, frequently asked questions on the class discussion forum.

Lars Satow (Satow, 2017) developed a model for the future, which describes the levels of learning facilitation by artificial intelligence teaching assistants:

- Level 1: Personalized messages from the teaching assistant welcomes new learners.
- Level 2: Teaching assistant advises learning materials, suggests following steps, possible collaborators and professionals for cooperative learning.
- Level 3: Teaching assistant responses to usual questions posted by students.
- Level 4: Teaching assistant establishes the steps to meet learning objectives and supervises the improvement of learning.
- Level 5: Teaching assistant gives personalized comments.
- Level 6: Teaching assistant offers individualized comments and endorsements, analyzes individual learning requests and provides tutoring instructions.

In this study, we analyzed 47 educational chatbots using the Facebook Messenger platform against the quality attributes of teaching, humanity, affect and accessibility and listed the top 10 educational chatbots.

With the exception of two (*Ask Frank, IFRSRookies*), most chatbots are mechanic in their behavior and answers, they lack basic interaction and communication patterns, their responses are based on button-based navigation, and they miss the skills of text-based command recognition. There was no significant difference in accessibility between educational chatbots and reactions were generally

weak. Only *Ask Frank* was able to respond to a simple questions, such as "How are you?", "Can we chat?" or "What is your name?".

Conversational style, tone, and attitude of the chatbot — the personality is critical to the success of natural dialogue between human and computer. Due to data provided by Facebook developers can easily personalize the greeting text using the person's name. We found that (N = 23, 49%) of educational chatbots were able to greet the user with a person's name. The Facebook Messenger platforms also support rich media, like images, animated GIFs and videos. These dynamic elements go a long way towards imbuing chatbots with personality. They make the conversation more enjoyable, more immersive, and more visually engaging. Emojis are another way to add personality in an entertaining and evocative way.

Based on the Lars Satow's model of learning facilitation, educational chatbots on Facebook Messenger platform vary between Level 1 and Level 4. Personalized welcome messages are simple to implement. Learning content is part of the chatbot knowledgebase, could be obtained from the web sources (e.g. a computational knowledge engine like Wolfram Alpha for *Ask Frank*), as well as providing links to websites. Chatbots are using pattern matching to classify text and produce a response. To be able to answer typical questions asked by learners, the intelligence of a chatbot depends on how these predefined patterns are defined and how well the text is analyzed and processed. Another way is to use a generative model with support of a neural network. This self-learning chatbot is trained using a large number of previous conversations with the users. It is always ready to have a response but this could be random and not always make a sense. Implementing a Question & Answer system to a chatbot looks like a default setting, but it requires a critical amount of data and expertise to create a conversational flow. Findings demonstrated, however, that all of the top ten educational chatbots were using quizzes or a Question & Answer module, which can help during the learning process for users. Gamification is a great way to engage students. Supporting this, results showed a game element, like a point system, was applied in four chatbots (*Feed.Mind, Kuni, NELA, Mastermind Games Bot*) out of the top ten. A broadcast functionality in a chatbot is able to send users reminders and keep them engaged with the chatbot. Users of the *Wordsworth* chatbot can sign up for a daily lesson or take a fun quiz to test their vocabulary skills.

4.1. Limitations

The present research was bound by several limitations, pointing out the directions for future improvement. For chatbot developers and even users, discoverability is one of the biggest challenges. There is no single place, where can we find chatbots on Facebook Messenger. One limitation was that chatbots for our research were selected from the bot-listed website (Botlist.com) using the Messenger and Education category as a filter. It is important to bear in mind that the scope of this research simulates possible user's searching strategy. These chatbots are promoting themselves as educational, no matter what is the actual content. Therefore, they come into sight of users using common search terms. Furthermore, the findings of this research provide insights for the limited view of publicly accessible and user-marketed educational chatbots from a time period of March–April 2018. We may obtain different results of the equivalent study when repeated at a different time.

5. Conclusions

The study examined the educational chatbots in the instant messaging application Facebook Messenger with focus to identify discoverability and characteristics as language, subject matter and developer's platform. The study also focused on the evaluation of selected chatbots to describe their current stage of learning facilitation using artificial intelligence. Educational chatbots on the Facebook Messenger platform vary from the basic level of sending personalized messages, to recommending learning content. These findings provide further evidence to suggest that chatbot programming (especially on the Facebook Messenger) is still in its early stages. Further features include artificial intelligence teaching assistants answering typical questions posted by learners, setting learning goals, and monitoring learning progress. The future research can be divided into two fields. The first field is to focus on the developers support to create and offer tools that allow any teacher to integrate chatbots into their classes without difficulty, and provide educational chatbot guidelines to successfully support coaching methods and students' learning. The second field is a content analysis of the actual conversations with students. It is technically possible to store, collect and analyze conversations from both macro and micro angles.

We hope this review will promote further research among other scholars similarly interested in using chatbots on the mobile platforms for teaching and learning purposes.

Acknowledgment

This research has been elaborated in the framework of the project No: SP2020/57 "Research and Development of Advanced Methods in the Area of Machines and Process Control" supported by the Ministry of Education, Youth and Sports, Czechia.

References

Awidi, I. T., Paynter, M., & Vujosevic, T. (2019). Facebook group in the learning design of a higher education course: An analysis of factors influencing positive learning experience for students. *Computers & Education, 129*, 106–121. https://doi.org/10.1016/j.compedu.2018.10.018.

Budiu, R. (2018). The user experience of chatbots, 11 25Retrieved from Nielsen Norman Group: https://www.nngroup.com/articles/chatbots/.

Chartrand, R. (2016). *Advantages and disadvantages of using mobile devices in a university language classroom*, 3 31 (Vol. 23). Bulletin of the Institute of Foreign Language Education Kurume University. Retrieved from http://hdl.handle.net/11316/445.

Cinglevue. (2017). *Learning and educational applications of chatbot technologies*, 8 11. Retrieved from Cinglevue: https://www.cinglevue.com/learning-educational-applications-chatbot-technologies/.

Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text & Talk, 34*(5), 545–567. https://doi.org/10.1515/text-2014-0018.
Constine, J. (2018). 2.5 billion people use at least one of Facebook's apps, 7 26Retrieved 7 26, 2018, from https://techcrunch.com/2018/07/25/facebook-2-5-billion-people/.
Crutzen, R., Peters, G.-J. Y., Portugal, S. D., Fisser, E. M., & Grolleman, J. J. (2011). An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study. *Journal of Adolescent Health, 48*, 514–519. https://doi.org/10.1016/j.jadohealth.2010.09.002.
Eicher, B., Polepeddi, L., & Goel, A. (2018). Jill Watson doesn't care if you're pregnant: Grounding AI ethics in empirical studies. In *Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society* (pp. 88–94). New York, NY, USA: ACM. https://doi.org/10.1145/3278721.3278760.
Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *The Internet and Higher Education, 19*, 18–26. https://doi.org/10.1016/j.iheduc.2013.06.002.
Gulz, A., Haake, M., Silvervarg, A., Sjödén, B., & Veletsianos, G. (2011). *Building a social conversational pedagogical agent: Design challenges and methodological approaches*. Hershey: IGI Global. https://doi.org/10.4018/978-1-60960-617-6.ch006.
Hartmans, A. (2017). These are the 10 most used smartphone apps, 8 29Retrieved from Business Insider https://www.businessinsider.com/most-used-smartphone-apps-2017-8.
Heller, B., Proctor, M., Mah, D., Jewell, L., & Cheung, B. (2005). Freudbot: An investigation of chatbot technology in distance education. In P. Kommers, & G. Richards (Eds.), *Proceedings of EdMedia + innovate learning 2005* (Vol. 6, pp. 3913–3918). Montreal, Canada: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/p/20691.
Hutchens, J. (1997). *How to pass the turing test by cheating*. University of Western Australia.
Johnson, K. (2018). Facebook Messenger passes 300,000 bots, 5 1Retrieved 7 26, 2018, from https://venturebeat.com/2018/05/01/facebook-messenger-passes-300000-bots/.
King, F. B. (2002). A virtual student: Not an ordinary Joe, 4 21 *The Internet and Higher Education, 5*(2), 157–166. https://doi.org/10.1016/S1096-7516(02)00085-4.
Laurillard, D. (2013). *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. London, UK: Routledge.
Laven, S. (1996). Claude - by brian mclaughlin. Retrieved from The Simon Laven Page: https://www.simonlaven.com/claude.htm.
Pereira, J., & Díaz, O. (2018). A quality analysis of Facebook messenger's most popular chatbots. In *Proceedings of the 33rd annual ACM symposium on applied computing* (pp. 2144–2150). New York, NY, USA: ACM. https://doi.org/10.1145/3167132.3167362.
Pimmer, C., Brühlmann, F., Odetola, T. D., Oluwasola, D. O., Dipeolu, O., & Ajuwon, A. J. (2019). Facilitating professional mobile learning communities with instant messaging. *Computers & Education, 128*, 102–112. https://doi.org/10.1016/j.compedu.2018.09.005.
R Core Team. (2014). *R: A language and environment for statistical*. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.
Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents, 04 15arXiv:1704.04579 https://arxiv.org/abs/1704.04579.
Roblyer, M. D., McDaniel, M., Webb, M., Herman, J., & Witty, J. V. (2010). Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites. *The Internet and Higher Education, 13*, 134–140. https://doi.org/10.1016/j.iheduc.2010.03.002.
Saaty, T. L. (1980). *The analytic hierarchy process*. New York, USA: McGraw-Hill.
Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research, 48*, 9–26. https://doi.org/10.1016/0377-2217(90)90057-I.
Saaty, T. L. (2008). Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process, 9 01 *RACSAM - Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas, 102*, 251–318. https://doi.org/10.1007/BF03191825.
Satow, L. (2017). Chatbots as teaching assistants: Introducing a model for learning facilitation by AI bots, 07 12Retrieved from https://blogs.sap.com/2017/07/12/chatbots-as-teaching-assistants-introducing-a-model-for-learning-facilitation-by-ai-bots/.
Schmulian, A., & Coetzee, S. A. (n.d.). The development of Messenger bots for teaching and learning and accounting students' experience of the use thereof. British Journal of Educational Technology, 0. doi:10.1111/bjet.12723.
Shawar, B. A., & Atwell, E. (2015). ALICE chatbot: Trials and outputs. *Computación Y Sistemas, 19*, 625–632. https://doi.org/10.13053/CyS-19-4-2326.
So, S. (2016). Mobile instant messaging support for teaching and learning in higher education. *The Internet and Higher Education, 31*, 32–42. https://doi.org/10.1016/j.iheduc.2016.06.001.
Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education, 94*, 252–275. https://doi.org/10.1016/j.compedu.2015.11.008.
Tallyn, E., Fried, H., Gianni, R., Isard, A., & Speed, C. (2018). The ethnobot: Gathering ethnographies in the age of IoT. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 604:1–604:13). New York, NY, USA: ACM. https://doi.org/10.1145/3173574.3174178.
Weinberg, C. (2017). How messenger and "M" are shifting gears, 02 22Retrieved 07 30, 2018, from https://www.theinformation.com/articles/how-messenger-and-m-are-shifting-gears.
Weizenbaum, J. (1983). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 26*, 23–28. https://doi.org/10.1145/357980.357991.
Westerman, D., Daniel, E. S., & Bowman, N. D. (2016). Learned risks and experienced rewards: Exploring the potential sources of students' attitudes toward social media and face-to-face communication. *The Internet and Higher Education, 31*, 52–57. https://doi.org/10.1016/j.iheduc.2016.06.004.
Wu, Y., Wang, G., Li, W., & Li, Z. (2008). Automatic chatbot knowledge acquisition from online forum via rough set and ensemble learning, 10. In *2008 IFIP international conference on network and parallel computing* (pp. 242–246). https://doi.org/10.1109/NPC.2008.24.
Yildirim, I. (2017). The effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. *The Internet and Higher Education, 33*, 86–92. https://doi.org/10.1016/j.iheduc.2017.02.002.